# Industry Organization and Output Size Distribution of Cotton Gins in the U.S.

Maria Mutuc
Cotton Economics Research Institute
Department of Agricultural and Applied Economics
Texas Tech University
Lubbock, TX 79409-2132
Email: maria.mutuc@ttu.edu

Darren Hudson
Cotton Economic Research Institute
Department of Agricultural and Applied Economics
Texas Tech University
Lubbock, TX 79409-2132
Email: darren.hudson@ttu.edu

***Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Orlando, FL, February 6-9, 2010***

# Industry Organization and Output Size Distribution of Cotton Gins in the U.S.

ABSTRACT

With cotton output declining by 46 percent from 2005-2008 (from 23.89 M bales in 2005 to 12.8 M bales in 2008), gins are processing less cotton. This paper examines how output size distribution of cotton gins in the U.S. has evolved and the extent to which the developments in the U.S. ethanol industry, specifically the passage of the Energy Policy Act in 2005 (and its subsequent revisions), have influenced this structural process. Markov transitional probability matrices (TPMs) are estimated for two periods: 1994-2004 and 2005-2008 to determine changes in output size distribution of gins. TPMs indicate that relative to the pre-2005 period, gins had a greater propensity to process lower outputs after 2005. It is purported that in industries constrained by declining demand, bigger firms with excess capacity operate at higher costs than smaller firms that operate closer to their minimum efficient scale.

*Keywords:* Cotton, cotton gins, transitional probabilities, Markov, minimum efficient scale

JEL Classification: Q10, Q12, L11

**Industry Organization and Output Size Distribution of Cotton Gins in the U.S.**[1]

## 1. Introduction

The United States cotton industry has not seen sustained acreage reductions since 1990 until recently. Unlike in the past where a contraction in area harvested in a given year was followed immediately by a recovery, recent experience has been different. In 2006, area harvested declined by 7.8% followed by successive reductions of 17.6% and 27.8% in 2007 and 2008 (Table 1). In 2009, however, acreage expanded slightly by 2.2%. These reductions in acreage coincided with the developments in the oil-corn-ethanol complex.

With the surge in crude oil prices from US$60 per barrel in 2005 to US$128 per barrel in 2008, the demand for ethanol as a fuel alternative significantly strengthened. This, in turn, expanded the demand for materials such as corn, among others, from which ethanol and other similar biofuels are created. In the U.S., as the ethanol industry absorbed a significant share of the corn crop, corn prices rose in recent years. Higher corn prices have provided farmers the incentive to switch acreage from competing crops to corn. One of these competing crops is cotton, the acreage for which has declined by as much as 45% from 2005 to 2008 (from 5.6 to 3.1 million hectares), the period following the passage of the Energy Policy Act of 2005 that mandates a new Renewable Fuel Standard (RFS). The RFS ensures that gasoline marketed in the U.S. contains a specific amount of renewable fuel. As a result, it is expected that between 2006 and 2012, the RFS is slated to increase demand for renewable fuels from 4.0 to 7.5 billion gallons per year (Baker and Zahnister, 2006). These mandates were subsequently expanded in 2007. This paper evaluates how the contemporaneous, recent declines in cotton acreage have affected the structure and costs faced by cotton gins in the U.S.

[1] The authors would like to acknowledge Brian Miller, a graduate student of the Department of Mathematics and Statistics at Texas Tech University for his invaluable assistance in mathematical programming.

## 2. Cotton Acreage and Output Size of Cotton Gins in the U.S.

Cotton production has historically tracked acreage movements and, as such, a similar trend can be observed in output as in acreage. From 1990 to 2004, cotton production increased steadily from 15.5 million bales to 23.2 million bales. In 2005, production increased to 23.9 million bales and started to decline until it reached 12.8 million bales in 2008. As a consequence, the average number of bales processed per gin in the U.S. declined from 26,920 bales in 2005 to 17,453 in 2008 (Figure 1). While cotton production and acreage have tracked each other closely, the number of gins has steadily declined for the past three decades (Figure 2). Even in the lead up to the period of increased cotton output prior to 2005, there were fewer gins every year, some of which have consolidated their operations and have existed alongside smaller gins.

To best illustrate what happened to the output size distribution of U.S. gins pre- and post-2005 (break that coincides with decline in cotton acreage), we derive transitional probability matrices (TPMs) for these two corresponding periods using maximum entropy econometrics for ill-posed problems developed by Golan and Amos (2001) implemented in Maple 13. The methodology and data used are discussed in the Appendix.[2] TPMs define the likelihood that a gin will move from producing at a particular output level at time $t$ (rows) to another level at time $t+1$ (columns) such that the diagonal elements of the matrix represents the likelihood that a gin will remain or continue to produce at the same level of output at time $t+1$ as in time $t$. For example, the $(i,j)$ entry represents the likelihood that a gin in the $i$th output category at time $t$ will move to the $j$th output category at time $t+1$. These transitional probabilities are derived using

[2] The complete program is available from the authors upon request.

maximum entropy econometrics for ill-posed problems developed by Golan and Amos (2001) implemented in *Maple 13*. This is an ill-posed problem because there are more unknowns than given values.

A cursory look at transitional probabilities across output sizes (probabilities or likelihood associated with the proportion of gins moving up and down output levels from one period to the next) reveals that there was a higher tendency for gins to move to higher production levels from 1994 to 2004 than there was for 2005 to 2008 (as seen from a comparison of output column with less than 40,000 in Tables 2a and 2b). In fact, the opposite was observed post-2005. There was a higher concentration of smaller gins that processed less than 15,000 bales and a lower concentration of gins that processed beyond 40,000 bales. The decline in acreage beginning 2005 has forced gins to process smaller volumes. To underscore this point, Tables 2a and 2b show that the probabilities along the upper triangle (which indicate the probabilities of gins moving to process higher output levels) are generally smaller in magnitude for 2005 to 2008 relative to pre-2005. This implies that gins, before 2005, were more likely to move up to higher output levels of production than for the period 2005-2008.

## 3. Theoretical Underpinnings: Firm Size and Minimum Efficient Scale (MES)

What is happening to the ginning industry in the U.S. is best illustrated using Figure 3. Figure 3 shows three average cost (AC) curves for three types of firms: small, medium, and large. Consider the *AC* curves for the medium and large firms. For the medium-sized firm, the output level at which its *AC* is minimum is $P_M$ at output $MES_M$ (minimum efficient scale for the medium-sized firm). Hence, if the medium-sized firm produces beyond this point, its *AC* starts to rise. For the large firm, its minimum efficient scale is at $MES_L$ that corresponds to price, $P_L$.

Again, should the large firm produce above this level, the AC it faces starts to rise. Notice that $MES_L > MES_M$ as the large firm benefits from economies of scale over a wider range of output. That is, the large firm (with higher fixed costs) is able to spread its fixed costs over a larger amount of output before it reaches a point near capacity when more than a proportional amount of the variable inputs are necessary to increase output compared to the medium-sized firm. To determine which kind of firms will operate in a particular industry, we have to take into the account the relative position of effective market demand with firms' average costs.

Even if large capacity firms are willing and able to operate at higher output levels, they are constrained by the effective market demand. In Figure 3, if the effective market demand's location shifts from a higher level to a lower level, say $Y^*$, two things can be observed: (a) it is more cost efficient for the medium-sized firm to produce at $Y^*$ relative to the large firm even if both can technically produce at $Y^*$, and (b) the average cost of production increases with the move to the lower output.

## 4. Empirical Application & Results

Using the average cost curves estimated by McPeek (1997) across four gin size categories for the ginning industry in the Texas Southern High Plains, the effects of acreage reduction on industry structure and average cost of ginning in the U.S. are empirically illustrated. McPeek classified gins according to their rated capacity: (a) 14 bph, (b) 21 bph, (c) 28 bph, and (d) 35 bph. All gins were assumed to operate for 19 hours per day, and for 71 days per season. Each of the four cost curves were estimated by McPeek (1997) using a computerized cost simulation program called GINMODEL. GINMODEL calculates the cost of ginning using both technical (engineering) and economic relationships derived from personal interviews with ginners and equipment manufacturers. McPeek's inputs to GINMODEL included input costs, investment costs, interest costs, depreciation and other relevant cost data. Without altering the relative average cost relationships across gin sizes, we updated McPeek's values. We added $0.25 to average fixed costs (to account for interest rate changes and inflation) and $0.50 to variable costs (to account for higher cost of bagging and ties). The average cost relationships used are shown in Table 3.

Based on Table 3, we computed for the resulting average cost of ginning per gin size across different output levels to find the minimum efficient scale for each gin size (assuming each gin operates for 19 hours per day for 71 days per season). The results are shown in Table 4. When the ginning industry averaged 26,319 bales per gin in 2004, it was profitable for size 2 gins to operate. However, when ginning volume declined in 2008 to 17,453 bales per gin, it was more cost efficient for size 1 gins to operate. Size 2 gins that continue to operate incurred more costs than size 1 gins. Using some interpolation between discrete cost points, in 2004, average total ginning cost was at $44.9 per bale while in 2008 it increased to $54.2 per bale. In relative terms, this increase in average costs represents about 17% of the average ginning cost in 2008.

The costs to the economy of the recent acreage reductions come in the form of higher ginning costs as well as costs sunk in fixed investments in the form of equipment and other fixtures made by larger gins that are likely to disinvest (cut capacity or close entirely) from industry. These costs need to be accounted for in looking at the policy effects of increased biofuel production in the U.S.

## 5. Conclusions & Extensions

As a result of sustained acreage reductions since 2006, the cotton ginning industry in the U.S. has seen the contraction of effective ginning demand to a level that, relative to firms' minimum efficient scale (MES), makes smaller gins more cost efficient (and capable of staking it out in the industry) given the smaller fixed costs and stranded investments attendant to smaller operations. Whereas in 2004, a typical gin processed 26,319 bales, this volume went down to only 17,453 bales in 2008. As a result, the industry is beginning to move towards smaller-sized and lesser number of gins. Also, concurrent with smaller volumes of output, the average cost of ginning has increased from $44.9 per bale in 2004 to $54.2 per bale in 2008. In relative terms, this increase in average costs represents about 17% of the average ginning cost in 2008.

The results from this study can be used to determine whether or not the cotton ginning industry will continue to shrink so that resources and policies can be allocated to avert or sustain the likelihood of such an outcome. Corresponding losses to the cotton industry can then be calculated and used for further analysis of welfare studies and income stabilization policies for cotton farmers.

The analysis can be refined to include, in the TPMs, categories for entry and exit to further clarify which types of gins are entering or leaving the industry altogether. This will enhance the discussion of output size distribution across gins. Also, analysis can be supplemented by investigating the contribution of different factors such as cotton prices, ownership type, among others, on the observed changes in transition probabilities.

## References


Baker, A., and S. Zahnister (2006). "Ethanol Reshapes the Corn Market." *Amber Waves*, April 2006.

Golan, A., G. Judge, D. Miller (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley and Sons, New York.

Maple 13 (2009). Maplesoft Company, Ontario, Canada.

McPeek, B. (1997). "Optimum Organization of the Cotton Ginning Industry in the Texas Southern High Plains." Unpublished Master's Thesis, Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, TX.

National Agricultural Statistics Service. Cotton Ginnings Annual Survey. Various issues available at http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1041

Varian, H. (1992). *Microeconomic Analysis.* Third Edition, W.W. Norton and Company, Inc.,

**Table 1.** Cotton Area Harvested for the United States, 1990-2009

| Marketing Year | Area Harvested (1000 hectares) | Growth Rate (%) |
|---|---|---|
| 1990/91 | 4748 | 23.01 |
| 1991/92 | 5245 | 10.47 |
| 1992/93 | 4501 | (14.18) |
| 1993/94 | 5173 | 14.93 |
| 1994/95 | 5391 | 4.21 |
| 1995/96 | 6478 | 20.16 |
| 1996/97 | 5216 | (19.48) |
| 1997/98 | 5425 | 4.01 |
| 1998/99 | 4324 | (20.29) |
| 1999/00 | 5433 | 25.65 |
| 2000/01 | 5282 | (2.78) |
| 2001/02 | 5596 | 5.94 |
| 2002/03 | 5025 | (10.20) |
| 2003/04 | 4858 | (3.32) |
| 2004/05 | 5284 | 8.77 |
| 2005/06 | 5586 | 5.72 |
| 2006/07 | 5152 | (7.77) |
| 2007/08 | 4245 | (17.60) |
| 2008/09 | 3063 | (27.84) |
| 2009/10 | 3129 | 2.15 |
| 1990-2004 | | 11.29 |
| 2005-2009 | | (43.98) |

Source: United States Department of Agriculture

**Table 2a.** Transitional Probability Matrix for the United States, 1994-2004

| | 2004 | | | |
|---|---|---|---|---|
| **1994** | <15,000 | 15,000-19,000 | 20,000-39,000 | >40,000 |
| <15,000 | 0.4002 | 0.0594 | 0.3386 | 0.2018 |
| 15,000-19,000 | 0.2865 | 0.1899 | 0.2764 | 0.2472 |
| 20,000-39,000 | 0.3182 | 0.1448 | 0.2970 | 0.2399 |
| >40,000 | 0.2712 | 0.2140 | 0.2656 | 0.2491 |

**Table 2b.** Transitional Probability Matrix for the United States, 2005-2008

| | 2008 | | | |
|---|---|---|---|---|
| **2005** | <15,000 | 15,000-19,000 | 20,000-39,000 | >40,000 |
| <15,000 | 0.6098 | 0.1013 | 0.2029 | 0.0859 |
| 15,000-19,000 | 0.3861 | 0.1886 | 0.2488 | 0.1766 |
| 20,000-39,000 | 0.5998 | 0.1049 | 0.2059 | 0.0894 |
| >40,000 | 0.5235 | 0.1331 | 0.2260 | 0.1174 |

**Table 3.** Average Cost Functions

| Rated Capacity (bales per hour) | Average Cost Function (US$/bale) |
|---|---|
| 14 | 25.04 + 508,214.93 * (1/number of bales) |
| 21 | 22.31 + 593,836.74 * (1/number of bales) |
| 28 | 22.14 + 623,010.17 * (1/number of bales) |
| 35 | 20.75 + 680,587.09 * (1/number of bales) |

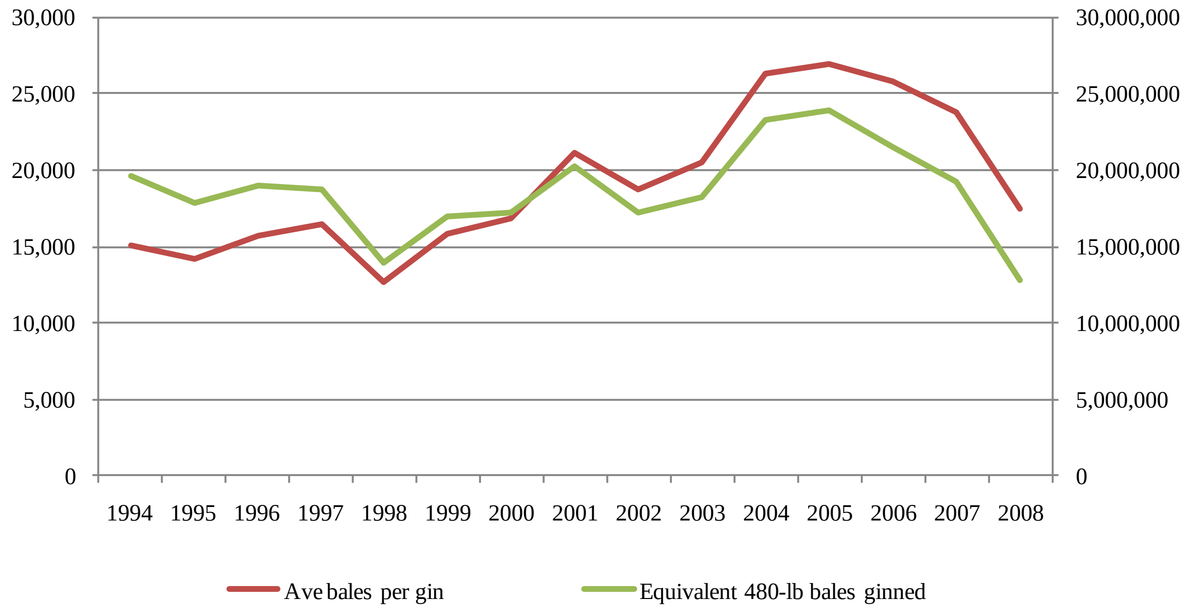
Source of basic data: McPeek (1997), with authors' calculations.

**Table 4.** Comparison of Average Cost Across Gin Sizes Across Output Levels

| Ginning rate (bales per hour) | Ginning volume (bales per season) | Average total cost (US$/bale) | | | |
|---|---|---|---|---|---|
| | | Size 1 **14 bph** | Size 2 **21 bph** | Size 3 **28 bph** | Size 4 **35 bph** |
| 6 | 8,094 | 87.83 | 95.67 | 99.12 | 104.84 |
| 7 | 9,443 | 78.86 | 85.19 | 88.12 | 92.82 |
| 8 | 10,792 | 72.13 | 77.33 | 79.87 | 83.82 |
| 9 | 12,141 | 66.90 | 71.22 | 73.46 | 76.81 |
| 10 | 13,490 | 62.71 | 66.33 | 68.33 | 71.20 |
| 11 | 14,839 | 59.29 | 62.33 | 64.13 | 66.62 |
| 12 | 16,188 | 56.43 | 58.99 | 60.63 | 62.79 |
| 13 | 17,537 | 54.02 | 56.17 | 57.67 | 59.56 |
| **14** | **18,886** | **51.95** | **53.75** | **55.13** | **56.79** |
| 15 | 20,235 | | 51.65 | 52.93 | 54.39 |
| 16 | 21,584 | | 49.82 | 51.01 | 52.28 |
| 17 | 22,933 | | 48.20 | 49.31 | 50.43 |
| 18 | 24,282 | | 46.76 | 47.80 | 48.78 |
| 19 | 25,631 | | 45.48 | 46.45 | 47.30 |
| 20 | 26,980 | | 44.32 | 45.23 | 45.98 |
| **21** | **28,329** | | **43.27** | **44.14** | **44.78** |
| 22 | 29,678 | | | 43.14 | 43.68 |
| 23 | 31,027 | | | 42.22 | 42.69 |
| 24 | 32,376 | | | 41.39 | 41.77 |
| 25 | 33,725 | | | 40.62 | 40.93 |
| 26 | 35,074 | | | 39.91 | 40.16 |
| 27 | 36,423 | | | 39.25 | 39.44 |
| **28** | **37,772** | | | **38.64** | **38.77** |
| 29 | 39,121 | | | | 38.15 |
| 30 | 40,470 | | | | 37.57 |
| 31 | 41,819 | | | | 37.03 |
| 32 | 43,168 | | | | 36.52 |
| 33 | 44,517 | | | | 36.04 |
| 34 | 45,866 | | | | 35.59 |
| **35** | **47,215** | | | | **35.17** |

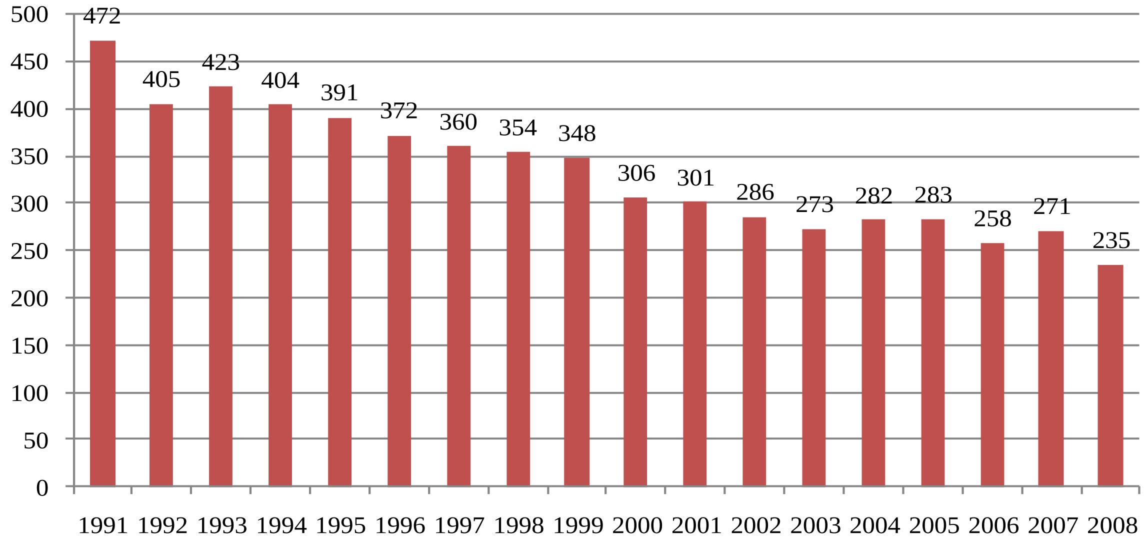
Source: Authors' computations

**Figure 1.** Total Bales Ginned in the U.S. and Average Bales Processed Per Gin



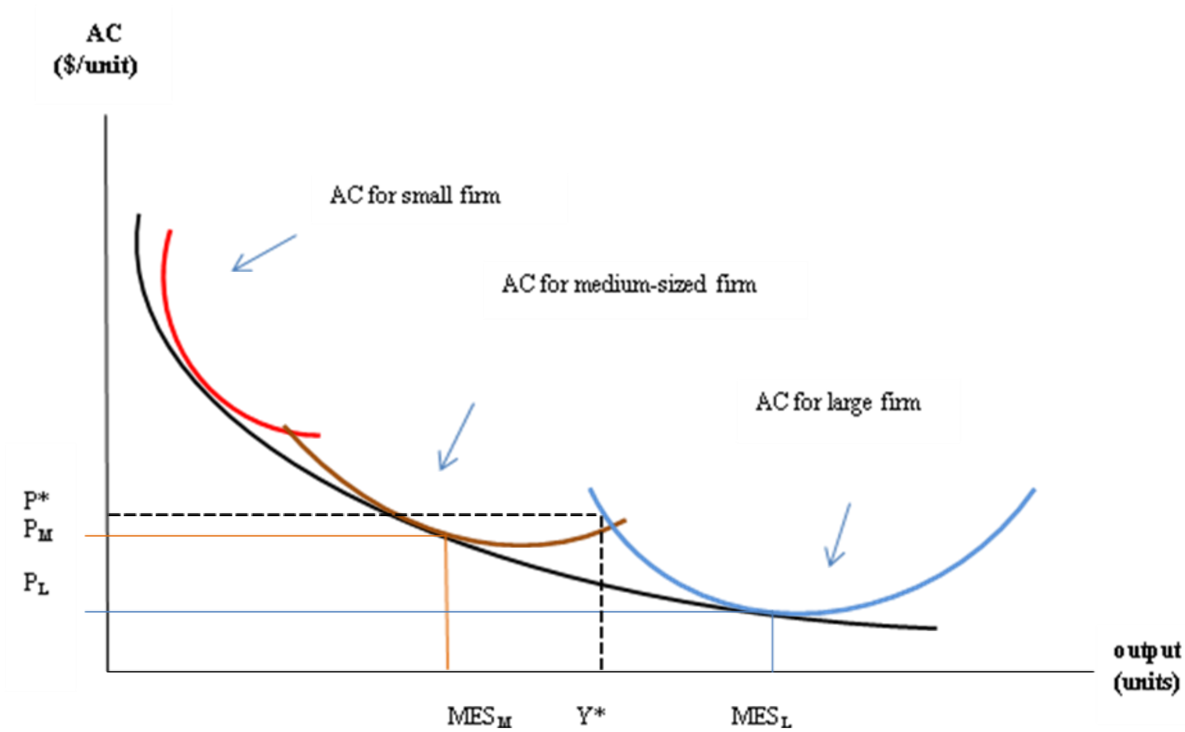

Source: NASS

**Figure 2.** Number of Cotton Gins in the United States, 1991-2008




Source: NASS

**Figure 3.** Average Cost and Minimum Efficient Scale



**Appendix:** Estimating Markov Transition Matrices Using Proportions Data

A.1. Data

Data used in this research come from the National Agricultural Statistics Service (NASS) on the number of gins and bales ginned by size group for the entire U.S. and on a state-level. They are published in the *Cotton Ginnings Summary*, from which data for years 1994-2008 were extracted. In the original data, output sizes (number of bales per year) are divided into 8 categories: (a) 1,000-2,999; (b) 3,000-4,999; (c) 5,000-6,999; (d) 7,000-9,999; (e) 10,000-14,999; (f) 15,000-19,999; (g) 20,000-39,999; and (h) 40,000 and over. The number of gins that fall under each category is recorded. From the original data, categories (a)-(e) were combined that reduces the total number of categories to 4: (a) less than 15,000; (b) 15,000-19,999; (c) 20,000-39,999; and (d) 40,000 and over. From these proportions were derived.

A.2. Maximum Entropy Basis for Information Recovery

The equations and methodology described here are from Golan, Judge & Miller (1996) as these exact equations were implemented in a technical computing software called *Maple 13* to derive transitional probability matrices using proportions data.

Given aggregate data, let the vector $\mathbf{x}(t)$ represent the ($K$ x 1) vector of proportion of gins falling in the $k$th Markov state (category) in time $t$, and $y(t + 1)$ represent the ($K$ x 1) vector of proportions of gins falling in each of the categories in time ($t$ + 1), then the stationary first-order Markov process may be written as

$$\mathbf{y}'(t+1) = \mathbf{x}'(t)P \qquad (1)$$

where $P = (\mathbf{p}_1 \quad \mathbf{p}_2 \ \ . . . \ \ \mathbf{p}_k)$ is an unknown and unobservable ($K$ x $K$) matrix of transition probabilities. If we rewrite the transition probabilities as $\mathbf{p} = (\mathbf{p}_1^{'}, \mathbf{p}_2^{'}, ..., \mathbf{p}_K^{'})'$, then we may rewrite (1) as

$$\underset{(K \times 1)}{\begin{bmatrix} y_1(t+1) \\ y_2(t+1) \\ \vdots \\ y_K(t+1) \end{bmatrix}} = \underset{(K \times K^2)}{\begin{bmatrix} \mathbf{x}'(t) & & & \\ & \mathbf{x}'(t) & & \\ & & \ddots & \\ & & & \mathbf{x}'(t) \end{bmatrix}} \underset{(K^2 \times 1)}{\begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \\ \mathbf{p}_K \end{bmatrix}} \tag{2}$$

or compactly for one transition as

$$\mathbf{y}_1 = X_1 \mathbf{p} \tag{3}$$

where $X_1 = I_K \otimes \mathbf{x}'(t)$ and $\otimes$ denotes the Kronecker product. If we let $T$ equal the number of data transition periods, then we may define our problem in terms of a vector of unknown transition probabilities that we wish to obtain from our data

$$\underset{(TK \times 1)}{\mathbf{y}_T} = \underset{(TK \times K^2)}{I_K \otimes X_T} \underset{(K^2 \times 1)}{\mathbf{p}} \tag{4}$$

If $TK < K^2$, the matrix $I_K \otimes X_T$ is non-invertible. In this case, traditional mathematical procedures yield solutions that contain $(K^2 - TK)$ arbitrary parameters.

To recover the transitional probabilities $\mathbf{p}$ for a $K$ state stationary Markov problem, when using data from $T$ transitions, we use the maximum entropy (ME) principle by Lee and Judge (1996) and state the problem as

$$\max_{p} \quad H(\mathbf{p}) = -\mathbf{p}'\ln\mathbf{p} = -\sum_i \sum_j p_{ij} \ln p_{ij} \tag{5}$$

subject to the first-order Markov condition

$$\underset{(TK \times K^2)}{I_K \otimes X_T} \ \mathbf{p} = \underset{(TK \times 1)}{\mathbf{y}_T} \tag{6}$$

the $K$ transition probability row sum constraints

$$\underset{(k \times k^2)}{\mathbf{1}' \otimes I_K} \ \mathbf{p} = \underset{(k \times 1)}{\mathbf{1}} \tag{7}$$

and

$$\mathbf{p} \geq \mathbf{0} \tag{8}$$

where $\mathbf{p}$ is a $(K^2 \times 1)$ vector, $\mathbf{1}$ is a $(K \times 1)$ vector of ones, $\mathbf{y}$ is a $(TK \times 1)$ vector of state outcomes for $T$ data transitions and $X_T$ is a $(TK \times K^2)$ matrix of state outcomes for $T$ transitions.

In scalar form, the corresponding Lagrangian equation is

$$\mathbf{L} = -\sum_i \sum_j p_{ij} \ln(p_{ij}) + \sum_i \sum_j \lambda_{tj} \left[ y_j(t+1) - \sum_i x_i(t) p_{ij} \right] + \sum_i \mu_i \left[ 1 - \sum_j p_{ij} \right] \tag{9}$$

and the optimal conditions are

$$\frac{\partial \mathbf{L}}{\partial p_{ij}} = -\ln(\hat{p}_{ij}) - 1 - \sum_t x_i(t)\hat{\lambda}_{tj} - \hat{\mu}_i = 0 \tag{10}$$

$$\frac{\partial \mathbf{L}}{\partial \lambda_{tj}} = y_j(t+1) - \sum_i x_i(t) p_{ij} = 0 \tag{11}$$

$$\frac{\partial \mathbf{L}}{\partial \mu_i} = 1 - \sum_j p_{ij} = 0 \tag{12}$$

Following this, the normalized ME solution is

$$\hat{p}_{ij} = \frac{\exp\left(-\sum_t x_i(t)\hat{\lambda}_{tj}\right)}{\sum_n \exp\left(-\sum_t x_i(t)\hat{\lambda}_{tn}\right)} = \frac{\exp\left(-\sum_t x_i(t)\hat{\lambda}_{tj}\right)}{\Omega_i(\hat{\lambda})} \tag{13}$$

and the dual or unconstrained objective is

$$M(\lambda) = \sum_t \sum_j y_j(t+1)\lambda_{tj} + \sum_i \ln(\Omega_i(\lambda)) \, . \tag{14}$$

Note that $T$ of the Lagrange multipliers are redundant; parameters can be normalized by setting $\hat{\lambda}_{t1} = 0$ for each $t = 1, \ldots, T$. We can arbitrarily scale each of the probabilities by dividing the numerator and denominator by the first numerator. This provides scaled solutions of the form

$$\hat{p}_{ij} = \begin{cases} \dfrac{1}{1 + \sum_{j=2}^{K} \exp\left(-\sum_t x_i(t)(\hat{\lambda}_{tj} - \hat{\lambda}_{t1})\right)}, & \text{for } j = 1 \\[2ex] \dfrac{\exp\left(-\sum_t x_i(t)(\hat{\lambda}_{tj} - \hat{\lambda}_{t1})\right)}{1 + \sum_{j=2}^{K} \exp\left(-\sum_t x_i(t)(\hat{\lambda}_{tj} - \hat{\lambda}_{t1})\right)}, & \textit{otherwise} \end{cases} \tag{15}$$

In addition, a multivariate steepest descent method is used to get a vector "close" to an answer.  That is, within some epsilon radius of a solution.  Although probabilities are not complex numbers, an algebraically closed field is necessary to solve all the equations.  To address this, the built-in function *fsolve* procedure in *Maple 13* is used over the complex numbers.